

# Spatially Varying Empirical Bayes Approach to Area-level models\*

SHONOSUKE SUGASAWA

*Risk Analysis Research Center, The Institute of Statistical Mathematics*

YUKI KAWAKUBO

*Faculty of Law, Politics and Economics, Chiba University*

KOTA OGASAWARA

*Department of Industrial Engineering and Economics, Tokyo Institute of Technology*

**Abstract.** We consider two-stage hierarchical area-level models based on natural exponentially family for small area estimation. While the areas are treated exchangeably and the model parameters are assumed to be the same over the areas, we might lose the efficiency when there exist spatial heterogeneity. To overcome this problem, we here propose two-stage area-level model with spatially varying model parameters and local marginal likelihood approach to estimating these parameters to compute empirical Bayes estimators of area means. We also discuss some related problems including mean squared error estimation, benchmarked estimation and estimation in non-sampled areas. The performance of the proposed method is evaluated through simulations and applications to two data sets.

**Key words:** empirical Bayes; hierarchical model; local likelihood; mean squared error; natural exponential family; small area estimation

## 1 Introduction

Small area estimation is widely used to produce reliable estimates of area (cluster) means with small, or even zero sample sizes. In the case of a small sample size, it is well-recognized that the direct estimator based only on the area-specific samples has the high variability and is not appropriate for practical use. Hence, we need to “borrow strength” from related areas and produce indirect (model-based) estimates of area-specific means. For the purpose, mixed modeling or hierarchical modeling approach has been used as a standard statistical tool. For comprehensive reviews of small area estimation techniques, see Pflaumer (2013) and Rao and Molina (2015).

In this paper, we focus on hierarchical area-level models for summarized area-level data. Let  $m$  be a number of areas,  $\{y_i, \mathbf{x}_i\}_{i=1, \dots, m}$  be the sampled data, where  $y_i$  is the direct estimator of a area mean  $\mu_i$ , satisfying  $E[y_i | \mu_i] = \mu_i$ , and  $\mathbf{x}_i$  is a vector of covariates associated with  $y_i$ . Typically,  $y_i$  is an unstable estimator of  $\mu_i$  in the sense that  $\text{Var}(y_i | \mu_i)$  is large due to the small sample size within the area, thereby we aim to increase the accuracy of estimating  $\mu_i$  by computing a model-based (indirect)

---

\*This version: January 24, 2017

estimator. To this end, we consider the area-level hierarchical model proposed by Ghosh and Maiti (2004), described as

$$\begin{aligned} f(y_i|\theta_i) &= \exp \{n_i(\theta_i y_i - \psi(\theta_i)) + c(y_i, n_i)\}, \\ \pi(\theta_i; \phi) &= \exp \{\nu(m_i \theta_i - \psi(\theta_i)) + C(\nu, m_i)\}, \end{aligned} \quad (1)$$

where  $m_i = \psi'(\mathbf{x}_i^t \boldsymbol{\beta})$  with  $\psi'(t) = d\psi/dt$ ,  $\theta_i$  is a natural parameter,  $n_i$  is a known scalar value typically related to the area sample size,  $\phi = (\boldsymbol{\beta}^t, \nu)^t$  is a vector of unknown model parameters common in all the areas,  $\psi(\cdot)$ ,  $c(\cdot, \cdot)$  and  $C(\cdot, \cdot)$  are functions specific to each distribution. Under the model, the area mean  $\mu_i$  is expressed as

$$\mu_i = E[y_i|\theta_i] = \psi'(\theta_i),$$

noting that  $E[\mu_i] = m_i$  under (1). Moreover, it is assumed that the conditional variance is a quadratic function of the conditional mean  $\mu_i$ , namely  $\text{Var}(y_i|\theta_i) = n_i^{-1}Q(\mu_i)$ , where  $Q(x) = v_0 + v_1x + v_2x^2$  for known constants  $v_0$ ,  $v_1$ , and  $v_2$ , which are not simultaneously zero. It is noted that  $\text{Var}(\mu_i) = Q(m_i)/(\nu - v_2)$  under this assumption. Then the distributions of  $y_i|\theta_i$  and  $\theta_i$  in (1) with the form of the conditional variance belong to the natural exponential family with quadratic variance function studied by Morris (1982, 1983). The model (1) is flexible enough to use in practice since the model includes well-used area-level models as special cases, Fay-Herriot model (Fay and Herriot, 1979), Poisson-gamma model (Clayton and Kaldor, 1987) and binomial-beta model (Williams, 1975). Hence, the formulation (1) enables us to deal with typical three models in the same framework.

It is remarked that the distribution of  $\theta_i$  in (1) is the conjugate prior. Then, the posterior distribution of  $\theta_i$  given  $y_i$  has the same form as  $\pi(\theta; \phi)$ , and the Bayes estimator or the conditional expectation of  $\mu_i$  can be obtained as

$$\tilde{\mu}_i \equiv \tilde{\mu}_i(y_i, \phi) = \frac{n_i y_i + \nu m_i}{n_i + \nu}, \quad (2)$$

which is the weighted mean of the direct estimator  $y_i$  and prior (synthetic) mean  $m_i$  with the weight determined by  $n_i$  and  $\nu$ . Moreover, owing to the conjugacy, the marginal distribution of  $y_i$  can be expressed as

$$f(y_i; \phi) = \exp \{c(y_i, n_i) + C(\nu, m_i) - C(n_i + \nu, \tilde{\mu}_i)\},$$

thereby the maximum likelihood estimator of  $\phi$  can be defined as the maximizer of the function with an analytical expression,  $\sum_{i=1}^m \log f(y_i; \phi)$ . By replacing the unknown parameter  $\phi$  with its maximum likelihood estimator, we easily get the empirical Bayes estimator  $\hat{\mu}_i \equiv \tilde{\mu}_i(y_i; \hat{\phi})$  of  $\mu_i$ . On the other hand, the generalized linear mixed models (Jiang, 2006) suffer from analytical tractability. Specifically, the posterior distribution of the area mean as well as the marginal distribution cannot be obtained in closed forms, so that we must rely on some computer intensive numerical method to compute the estimator of the area mean. Hence, the model (1) can be seen as an useful alternative to the generalized linear mixed models in area-level data analysis.

It should be pointed out that the model parameters in (1) are assumed to be the same over all the areas, which means that all the areas are treated exchangeably. However, one might lose the efficiency of estimating the area mean through the non-spatial model (1) if there is spatial heterogeneity. In the context of generalized linear mixed

model, the simultaneous autoregressive structures for spatially correlated normal random effects are often used in both normal and non-normal models, for example, in Bandyopadhyay et al. (2009), Marhuenda et al. (2013), Wakefield (2007). On the other hand, in the context of (1), the modeling spatial correlations among  $\theta_i$ 's is not straightforward. Instead, we assume that the model parameters  $\phi$  in (1) smoothly vary from areas to areas. Specifically, letting  $\mathbf{u}_i$  is the location of  $i$ th area, we allow  $\phi$  to depend on  $\mathbf{u}_i$ , namely  $\phi = \phi(\mathbf{u}_i)$ . For estimating the spatially varying parameter  $\phi(\mathbf{u}_i)$ , we apply the local likelihood method (Tibshirani and Hastie, 1987) to the marginal likelihood. Clearly, this method is closely related to the geographically weighted regression (Brunsdon et al., 1998; Fotheringham et al., 2002) in which the regression coefficients are assumed to be different in each location. In the context of small area estimation, the technique was used in Salvati et al. (2012) and Chandra et al. (2012) for continuous data, and Chambers et al (2014) for count data while the method is not based on hierarchical models but on quantile regressions.

The rest of the paper is organized as follows. In Section 2, we propose spatially varying empirical Bayes methods and discuss some related problems including mean squared error estimation, benchmarked estimation and estimation in non-sampled areas. In Section 3, we describe typical three models belonging to our framework. In Section 4, we evaluated the finite sample performances of the proposed methods through simulations. In Section 5, we show two results of applications to Scottish lip cancer data by Poisson-gamma model and Spanish poverty rate data by binomial-beta model. Finally, some discussions are given in Section 6.

## 2 Spatially Varying Empirical Bayes Methods

### 2.1 Spatially varying models and local likelihood estimation

We introduce a spatially varying area-level models by allowing the model parameters in (1) to vary spatially, which are described as

$$\begin{aligned} f(y_i|\theta_i) &= \exp \{n_i(\theta_i y_i - \psi(\theta_i)) + c(y_i, n_i)\}, \\ \pi(\theta_i; \phi(\mathbf{u}_i)) &= \exp \{\nu(\mathbf{u}_i)(m_i(\mathbf{u}_i)\theta_i - \psi(\theta_i)) + C(\nu(\mathbf{u}_i), m_i(\mathbf{u}_i))\}, \end{aligned} \quad (3)$$

where  $m_i(\mathbf{u}_i) = \psi'(\mathbf{x}_i^t \boldsymbol{\beta}(\mathbf{u}_i))$ ,  $\mathbf{u}_i = (u_{1i}, u_{2i})$  represents the coordinate of the  $i$ th area, and  $\phi(\mathbf{u}_i) = (\boldsymbol{\beta}(\mathbf{u}_i)^t, \nu(\mathbf{u}_i))^t$  is the spatially varying model parameters. It is noted that the first stage model of  $y_i|\theta_i$  is the same as (1) while the prior distribution of  $\theta_i$  is spatially varying.

To estimate the spatially varying parameters  $\phi(\mathbf{u}_i)$ , we use the local likelihood method (Tibshirani and Hastie, 1987) and suggest estimating  $\phi(\mathbf{u}_i)$  by maximizing the following locally weighted log-likelihood function:

$$\ell(\phi(\mathbf{u}_i)) = \sum_{k=1}^m w(\|\mathbf{u}_i - \mathbf{u}_k\|) \left\{ C(\nu(\mathbf{u}_i), m_k(\mathbf{u}_i)) - C(n_k + \nu(\mathbf{u}_i), \tilde{\mu}_k(y_k, \phi(\mathbf{u}_i))) \right\}, \quad (4)$$

where  $w(\cdot)$  is a user-specified kernel function and  $\tilde{\mu}_k(y_k, \phi(\mathbf{u}_i)) = (n_k + \nu(\mathbf{u}_i))^{-1}(n_k y_k + \nu(\mathbf{u}_i) m_k(\mathbf{u}_i))$  with  $m_k(\mathbf{u}_i) = \psi'(\mathbf{x}_k^t \boldsymbol{\beta}(\mathbf{u}_i))$ . The weights  $w(\|\mathbf{u}_i - \mathbf{u}_k\|)$  should gradually decrease as the distance between two locations  $\mathbf{u}_i$  and  $\mathbf{u}_k$  becomes larger. A common choice of the kernel  $w(\cdot)$  is the Gaussian kernel (Brunsdon et al., 1996) defined as

$$w(\|\mathbf{u}_i - \mathbf{u}_k\|) = \exp \left( -\frac{\|\mathbf{u}_i - \mathbf{u}_k\|^2}{2b^2} \right),$$

where  $b$  is the bandwidth controlling the rate at which the weight declines as the distance between two locations. The weight decays slowly under large  $b$  while the weight decays rapidly under small  $b$ .

As mentioned earlier, the local likelihood depends on the unknown bandwidth  $b$ . Following Fotheringham et al. (2002), we select the bandwidth  $b$  based on the cross-validation criterion given by

$$\text{CV}(b) = \sum_{i=1}^m \{y_i - \hat{y}_{(-i)}(b)\}^2, \quad (5)$$

where  $\hat{y}_{(-i)}(b)$  is the estimated value for location  $i$ , omitting the observation  $y_i$ . Specifically, we use the synthetic estimator  $\hat{y}_{(-i)}(b) = \psi'(\mathbf{x}_i^t \hat{\boldsymbol{\beta}}_{(-i)}(\mathbf{u}_i))$ , where  $\hat{\boldsymbol{\beta}}_{(-i)}(\mathbf{u}_i)$  is the estimated values from the local likelihood without  $y_i$ . For searching  $b$  minimizing  $\text{CV}(b)$ , we use golden section search (Brent, et al, 1973) over the interval  $[b_\ell, b_u]$ . Here  $b_\ell > 0$  should set to be small, for example  $b_\ell = 0.01$ , and we suggest setting  $b_u = 2 \max_{i,k} \|\mathbf{u}_i - \mathbf{u}_k\|^2$ .

Under the model (3), the Bayes estimator of  $\mu_i = E[y_i|\theta_i]$  is the same form as (2) and the empirical Bayes estimator is

$$\hat{\mu}_i^{\text{SV}} \equiv \tilde{\mu}_i(y_i, \hat{\boldsymbol{\phi}}(\mathbf{u}_i)) = \frac{n_i y_i + \hat{\nu}(\mathbf{u}_i) \hat{m}_i(\mathbf{u}_i)}{n_i + \hat{\nu}(\mathbf{u}_i)},$$

which we call spatially varying empirical Bayes (SVEB) estimator.

## 2.2 Hybrid bootstrap mean squared error estimation

In real applications, measuring uncertainty of the empirical Bayes estimator is required to assess the reliability of the estimates. Traditionally, an estimator of mean squared errors have been used for the purpose, see Prasad and Rao (1990) and Datta et al. (2005).

The MSE of the empirical Bayes estimator  $\hat{\mu}_i$  can be expressed as

$$\begin{aligned} \text{MSE}_i &= E[(\hat{\mu}_i - \mu_i)^2] = E[(\tilde{\mu}_i - \mu_i)^2] + E[(\hat{\mu}_i - \tilde{\mu}_i)^2] \\ &\equiv R_{1i}(\boldsymbol{\phi}(\mathbf{u}_i)) + R_{2i}(\boldsymbol{\phi}(\mathbf{u}_i)), \end{aligned}$$

since  $\tilde{\mu}_i = E[\mu_i|y_i]$ . Owing to the quadratic variance function, it follows that

$$R_{1i}(\boldsymbol{\phi}(\mathbf{u}_i)) = \frac{\nu(\mathbf{u}_i)Q(m_i(\mathbf{u}_i))}{(n_i + \nu(\mathbf{u}_i))(\nu(\mathbf{u}_i) - v_2)}.$$

Concerning the second term  $R_{2i}(\boldsymbol{\phi}(\mathbf{u}_i))$ , it vanishes as  $m \rightarrow \infty$ , thereby a naive (primitive) estimator of MSE is

$$\widehat{\text{MSE}}_i^N = R_{1i}(\hat{\boldsymbol{\phi}}(\mathbf{u}_i)). \quad (6)$$

However, if  $m$  is not sufficiently large,  $R_{2i}$  is not necessarily negligible, and the naive estimator could underestimate the true MSE. Moreover, the plug-in estimator  $R_{1i}(\hat{\boldsymbol{\phi}}(\mathbf{u}_i))$  is known to have a considerable bias.

To estimate the MSE more accurately than the naive MSE estimator (6), we use the hybrid bootstrap approach used in Butar and Lahiri (2003). Let  $\{y_1^s, \dots, y_m^s\}$  be the parametric bootstrap samples generated from the model (3) with  $\boldsymbol{\phi}(\mathbf{u}_i) = \hat{\boldsymbol{\phi}}(\mathbf{u}_i)$ ,

and define  $\hat{\phi}^s(\mathbf{u}_i)$  as the estimator computed from the bootstrap samples. Then the hybrid bootstrap MSE estimator is given by

$$\widehat{\text{MSE}}_i = 2R_{1i}(\hat{\phi}(\mathbf{u}_i)) - \frac{1}{B} \sum_{s=1}^B R_{1i}(\hat{\phi}^s(\mathbf{u}_i)) + \frac{1}{B} \sum_{s=1}^B \left\{ \tilde{\mu}_i(y_i^s, \hat{\phi}^s(\mathbf{u}_i)) - \tilde{\mu}_i(y_i^s, \hat{\phi}(\mathbf{u}_i)) \right\}^2. \quad (7)$$

Note that the last term corresponds to the estimator of  $R_{2i}$ , and first two terms correspond to a bias-corrected estimator of  $R_{1i}$ . Here we used an additive form for bias correction in estimating  $R_{1i}$ , other several forms have been proposed, see Hall and Maiti (2006).

### 2.3 Benchmarked estimation

The potential difficulty of an empirical Bayes estimator is that a (weighted) sum of empirical Bayes estimates is not necessarily equal to the corresponding direct estimates. Moreover, empirical Bayes estimates sometimes produce over-shrunk estimates, which results in inaccurate estimates of small area means. To avoid these problems, the benchmarked estimator (Datta et al., 2011; Bell et al, 2013) has been used as a standard tool in small area estimation. Here we consider the constraint  $\sum_{i=1}^m c_i \hat{\mu}_i = \sum_{i=1}^m c_i y_i$  with some known weight  $c_i$  satisfying  $\sum_{i=1}^m c_i = 1$ . A typical choice is  $c_i = n_i / \sum_{k=1}^m n_k$ . From Datta et al. (2011), the constrained empirical Bayes estimator  $\hat{\mu}_i^C$  that minimizes the squared error  $\sum_{i=1}^m \text{E}[(\hat{\mu}_i^C - \mu_i)^2]$  has the form

$$\hat{\mu}_i^C = \hat{\mu}_i + \omega_i \sum_{k=1}^m c_k (y_k - \hat{\mu}_k), \quad (8)$$

with  $\omega_i = c_i / \sum_{k=1}^m c_k^2$ . The weight  $c_i$  often satisfies  $\max_{1 \leq i \leq m} c_i = O(m^{-1})$  like  $c_i = n_i / \sum_{k=1}^m n_k$ . Then, the difference between  $\hat{\mu}_i^C$  and  $\hat{\mu}_i$  decreasing as the number of areas  $m$  gets large, namely, the differences are negligible when  $m$  is sufficiently large.

Since the benchmarked estimator increases the mean squared errors compared to the empirical Bayes estimator, we need to assess how large the excess MSE is. Regarding this issue, Steorts and Ghosh (2013) and Kubokawa et al. (2014) investigated the MSE estimators of benchmarked empirical Bayes estimators in area-level models via analytical or numerical way. Here, Similarly to Kubokawa et al. (2014), we consider a bootstrap method for evaluating the excess MSE. The excess MSE is expressed as

$$\begin{aligned} \text{EMSE}_i &= \text{E}[(\hat{\mu}_i^C - \mu_i)^2] - \text{E}[(\hat{\mu}_i - \mu_i)^2] \\ &= \text{E}[(\hat{\mu}_i^C - \hat{\mu}_i)^2] + 2\text{E}[(\hat{\mu}_i^C - \hat{\mu}_i)(\hat{\mu}_i - \tilde{\mu}_i)]. \end{aligned}$$

Therefore, the similar parametric bootstrap procedure used in the previous section enables us to estimate the excess MSE:

$$\widehat{\text{EMSE}}_i = \frac{1}{B} \sum_{s=1}^B (\hat{\mu}_i^{C,s} - \hat{\mu}_i^s)^2 + \frac{2}{B} \sum_{s=1}^B (\hat{\mu}_i^{C,s} - \hat{\mu}_i^s) \left\{ \hat{\mu}_i^s - \tilde{\mu}_i(y_i^s, \hat{\phi}(\mathbf{u}_i)) \right\}, \quad (9)$$

where  $\hat{\mu}_i^s = \tilde{\mu}_i(y_i^s, \hat{\phi}^s(\mathbf{u}_i))$  and  $\hat{\mu}_i^{C,s}$  is the benchmarked estimator (8) by replacing  $y_i$  and  $\hat{\mu}_i$  with  $y_i^s$  and  $\hat{\mu}_i^s$ , respectively.

## 2.4 Estimation in non-sampled area

In real applications, there could exist small areas with zero sample sizes. Let  $j$  be the index of a non-sampled area, and it is assumed that the covariate  $\mathbf{x}_j$  is available. In the traditional hierarchical model (1), the reasonable estimator of  $\mu_j$  is

$$\tilde{\mu}_j(\boldsymbol{\beta}) = m_j = \psi'(\mathbf{x}_j^t \boldsymbol{\beta}),$$

which is obtained by putting  $n_i = 0$  in the Bayes estimator (2). Similarly, we can define the estimator of  $\mu_j$  under the spatially varying model (3) as

$$\tilde{\mu}_j(\boldsymbol{\beta}(\mathbf{u}_j)) = m_j(\mathbf{u}_j) = \psi'(\mathbf{x}_j^t \boldsymbol{\beta}(\mathbf{u}_j)), \quad (10)$$

which is expected to provide more accurate estimates of  $\mu_j$  than the traditional one by using local information of non-sampled areas. The estimator of  $\boldsymbol{\beta}(\mathbf{u}_j)$  can be obtained by the local weighted likelihood (4) without the information in area  $j$ :

$$\ell(\boldsymbol{\phi}(\mathbf{u}_j)) = \sum_{k=1, k \neq j}^m w(\|\mathbf{u}_j - \mathbf{u}_k\|) \left\{ C(\nu(\mathbf{u}_j), m_k(\mathbf{u}_j)) - C(n_k + \nu(\mathbf{u}_j), \tilde{\mu}_k(y_k, \boldsymbol{\phi}(\mathbf{u}_j))) \right\}.$$

Thus the empirical version of (10) is given by  $\hat{m}_j(\mathbf{u}_j) = \psi'(\mathbf{x}_j^t \hat{\boldsymbol{\beta}}(\mathbf{u}_j))$ . The MSE of  $\hat{m}_j(\mathbf{u}_j)$  can be expressed as

$$\begin{aligned} \text{MSE}_j &= \text{E}[(\hat{m}_j(\mathbf{u}_j) - \mu_j)^2] = \text{E}[(\hat{m}_j(\mathbf{u}_j) - m_j + m_j - \mu_j)^2] \\ &= \frac{Q(m_i)}{\nu - v_2} + \text{E}[(\hat{m}_j(\mathbf{u}_j) - m_j)^2] + 2\text{E}[(\hat{m}_j(\mathbf{u}_j) - m_j)(m_j - \mu_j)], \end{aligned}$$

which can be estimated by the similar parametric bootstrap method given in Section 2.2.

## 3 Typical Models

### 3.1 Fay-Herriot model

When we assume that distributions of  $y_i|\theta_i$  and  $\theta_i$  are both normal with  $n_i = D_i^{-1}$ ,  $\nu = A^{-1}$ ,  $\psi(\theta_i) = \theta_i^2/2$ ,  $v_1 = v_2 = 0$  and  $v_0 = 1$ , the model (3) corresponds to the Fay-Herriot model (Fay and Herriot, 1979) with spatially varying parameters, described as

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta}(\mathbf{u}_i) + \sqrt{A(\mathbf{u}_i)} v_i + \sqrt{D_i} \varepsilon_i, \quad i = 1, \dots, m,$$

where  $v_i$ 's and  $\varepsilon_i$ 's are mutually independent standard normal random variables, and  $D_i$ 's are known sampling variances. Under the model, the marginal distribution of  $y_i$  is also normal,  $N(\mathbf{x}_i^t \boldsymbol{\beta}(\mathbf{u}_i), A(\mathbf{u}_i) + D_i)$ , thereby the Fisher-scoring algorithm is easily implemented for maximizing local likelihood (4). The algorithm entails updating the current estimates  $\boldsymbol{\beta}_i^{(r)}$  and  $A_i^{(r)}$  as

$$\begin{aligned} \boldsymbol{\beta}_i^{(r+1)} &= \boldsymbol{\beta}_i^{(r)} + \left( \sum_{k=1}^m \frac{w_{ik} \mathbf{x}_k \mathbf{x}_k^t}{A_i^{(r)} + D_k} \right)^{-1} \sum_{k=1}^m \frac{w_{ik} (y_k - \mathbf{x}_k^t \boldsymbol{\beta}_i^{(r)}) \mathbf{x}_k}{A_i^{(r)} + D_k} \\ A_i^{(r+1)} &= A_i^{(r)} - \left( \sum_{k=1}^m \frac{w_{ik}}{(A_i^{(r)} + D_k)^2} \right)^{-1} \sum_{k=1}^m w_{ik} \left( \frac{(y_k - \mathbf{x}_k^t \boldsymbol{\beta}_i^{(r)})^2}{(A_i^{(r)} + D_k)^2} - \frac{1}{A_i^{(r)} + D_k} \right), \end{aligned}$$

where  $w_{ik} = w(\|\mathbf{u}_i - \mathbf{u}_k\|)$ . This step is repeated until numerical convergence to get the estimates  $\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}}(\mathbf{u}_i)$  and  $\hat{A}_i = \hat{A}(\mathbf{u}_i)$ .

### 3.2 Poisson-gamma model

When the distributions of  $z_i (\equiv n_i y_i) | \lambda_i$  and  $\lambda_i \equiv \exp(\theta_i)$  are assumed to be Poisson and gamma, respectively, with  $\psi(\theta_i) = \exp(\theta_i)$ ,  $v_0 = v_2 = 0$  and  $v_1 = 1$ , the model (3) is expressed as

$$z_i | \lambda_i \sim \text{Po}(n_i \lambda_i) \quad \lambda_i \sim \Gamma(\nu(\mathbf{u}_i) m_i(\mathbf{u}_i), \nu(\mathbf{u}_i)), \quad i = 1, \dots, m.$$

where  $\lambda_1, \dots, \lambda_m$  are mutually independent,  $\text{Po}(\lambda)$  denotes the Poisson distribution with mean  $\lambda$ , and  $\Gamma(a, b)$  denotes the gamma distribution with density

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \quad x > 0.$$

The model corresponds to the Poisson-gamma model proposed in Clayton and Kaldor (1987) with spatially varying model parameters. It is well-known that the marginal distribution of  $z_i$  is the negative binomial distribution with probability function

$$f_m(z_i, \phi(\mathbf{u}_i)) = \frac{\Gamma(z_i + \nu(\mathbf{u}_i) m_i(\mathbf{u}_i))}{\Gamma(z_i + 1) \Gamma(\nu(\mathbf{u}_i) m_i(\mathbf{u}_i))} \left( \frac{n_i}{n_i + \nu(\mathbf{u}_i)} \right)^{n_i y_i} \left( \frac{\nu_i}{n_i + \nu(\mathbf{u}_i)} \right)^{\nu(\mathbf{u}_i) m_i(\mathbf{u}_i)},$$

and the local likelihood (4) is similar to the likelihood of geographical weighted negative binomial regression model suggested in Silva and Rodrigues (2014). For maximizing the weighted likelihood (4), we might use the iteratively reweighted least squares procedure proposed in Silva and Rodrigues (2014). On the other hand, owing to the conjugacy of the prior distribution of  $\lambda_i$ , the Expectation-Maximization (EM) algorithm (Dempster et al, 1977) is also an attractive tool. With the current estimates  $\beta_i^{(r)}$  and  $\nu_i^{(r)}$ , the algorithm entails computing  $E[\log u_i]$  with  $u_i \sim \Gamma(n_i y_i + \nu_i^{(r)} m_i^{(r)})$  and  $m_i^{(r)} = \exp(\mathbf{x}_i^t \beta_i^{(r)})$  as the E-step, and updating the current values by maximizing the weighted gamma likelihood with respect to  $\beta$  and  $\nu$ :

$$\sum_{i=1}^m w(\|\mathbf{u}_i - \mathbf{u}_k\|) \left\{ \nu m_i \log \nu - \log \Gamma(\nu m_i) + \nu m_i E[\log u_i] - \nu \frac{n_i y_i + \nu_i^{(r)} m_i^{(r)}}{n_i + \nu_i^{(r)}} \right\}.$$

### 3.3 Binomial-beta model

When the distributions of  $z_i (\equiv n_i y_i) | p_i$  and  $p_i \equiv \text{logistic}(\theta_i)$  with  $\text{logistic}(x) = \exp(x) / (1 + \exp(x))$  are assumed to be binomial and beta, respectively, with  $\psi(\theta_i) = \log(1 + \exp(\theta_i))$ ,  $v_0 = 0$ ,  $v_1 = 1$  and  $v_2 = -1$ , the model (3) is expressed as

$$z_i | p_i \sim \text{Bin}(n_i, p_i) \quad p_i \sim \text{Beta}(\nu(\mathbf{u}_i) m_i(\mathbf{u}_i), \nu(\mathbf{u}_i) (1 - m_i(\mathbf{u}_i))), \quad i = 1, \dots, m,$$

where  $\text{Beta}(a, b)$  denotes the beta distribution with density

$$f(x) = B(a, b)^{-1} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1,$$

and  $B(a, b)$  is the beta function. This model can be regarded as the extension of the binomial-beta model used in Williams (1975) in terms of the spatially varying hyperparameters. Under the model, the marginal probability function of  $z_i$  can be obtained as

$$f_m(z_i, \phi(\mathbf{u}_i)) = \binom{n_i}{z_i} \frac{B(z_i + \nu(\mathbf{u}_i) m_i(\mathbf{u}_i), n_i - z_i + \nu(\mathbf{u}_i) (1 - m_i(\mathbf{u}_i)))}{B(\nu(\mathbf{u}_i) m_i(\mathbf{u}_i), \nu(\mathbf{u}_i) (1 - m_i(\mathbf{u}_i)))},$$



which is not a familiar form. For maximizing the weighted likelihood (4), we can use the EM algorithm owing to the conjugacy of the prior of  $p_i$ . Based on the current estimates  $\beta_i^{(r)}$  and  $\nu_i^{(r)}$ , we first compute the two expectations  $E[\log u_i]$  and  $E[\log(1 - u_i)]$  with  $u_i \sim \text{Beta}(z_i + \nu_i^{(r)} m_i^{(r)}, n_i - z_i + \nu_i^{(r)}(1 - m_i^{(r)}))$  and  $m_i^{(r)} = \text{legit}(\mathbf{x}_i^t \beta_i^{(r)})$ , and then update the current estimates by maximizing the weighted beta likelihood with respect to  $\beta$  and  $\nu$ :

$$\sum_{i=1}^m w(\|\mathbf{u}_i - \mathbf{u}_k\|) \{ \nu m_i E[\log u_i] + \nu(1 - m_i) E[\log(1 - u_i)] - \log B(\nu m_i, \nu(1 - m_i)) \}.$$

## 4 Simulation studies

### 4.1 Estimation error comparison in sampled area

We first investigate estimation errors of the proposed estimator with the traditional estimator in finite samples. We first considered the Poisson-gamma model described in Section 3. The coordinates  $\mathbf{u}_i = (u_{1i}, u_{2i})$  were generated from the uniform distribution on  $(0, 1) \times (0, 1)$ , and covariate  $x_i$  was generated from the uniform distribution on  $(-1, 1)$ . Then we generated the simulated data from the following Poisson-gamma model:

$$z_i | \mu_i \sim \text{Po}(n_i \mu_i), \quad \mu_i \sim \Gamma(\nu(\mathbf{u}_i) m_i(\mathbf{u}_i), \nu(\mathbf{u}_i)), \quad i = 1, \dots, m, \quad (11)$$

where  $m = 60$ ,  $n_i = 20$ ,  $\mathbf{u}_i = (u_{1i}, u_{2i})$  and  $m_i(\mathbf{u}_i) = \exp(\beta_0(\mathbf{u}_i) + \beta_1(\mathbf{u}_i)x_i)$ . For setting hyperparameters, we considered two scenarios: (I) spatially varying:  $\beta_0(\mathbf{u}_i) = u_{1i} - u_{2i} - 1$ ,  $\beta_1(\mathbf{u}_i) = \sqrt{u_{1i}^2 + u_{2i}^2}$  and  $\nu(\mathbf{u}_i) = n_i \exp(u_{1i} + u_{2i} - 1)$ , and (II) spatially constant:  $\beta_0(\mathbf{u}_i) = 0.1$ ,  $\beta_1(\mathbf{u}_i) = 0.7$  and  $\nu(\mathbf{u}_i) = 50$ . It should be noted that the non-spatial Poisson-gamma model is the true model in setting (II) and the proposed spatially varying method is over-specified in this case. We applied both spatially varying (SV) method and spatially constant (SC) method assuming  $\beta_0(\mathbf{u}_i) = \beta_0, \beta_1(\mathbf{u}_i) = \beta_1$  and  $\nu(\mathbf{u}_i) = \nu$ , to the simulated data and computed the empirical Bayes estimates  $\hat{\mu}_i$  of  $\mu_i$ . Based on  $R = 1000$  simulation runs, we simulated the area-level MSE defined as

$$\text{MSE}_i = \frac{1}{R} \sum_{r=1}^R \left( \hat{\mu}_i^{(r)} - \mu_i^{(r)} \right)^2, \quad (12)$$

where  $\hat{\mu}_i^{(r)}$  and  $\mu_i^{(r)}$  denotes the estimated and true values of  $\mu_i$  in the  $r$ th simulation run. To compare the results between SV and SC, we computed the percent relative difference of the root of MSE:

$$\text{RD}_i = \frac{\sqrt{\text{MSE}_i^{\text{SV}}} - \sqrt{\text{MSE}_i^{\text{SC}}}}{\sqrt{\text{MSE}_i^{\text{SC}}}} \times 100, \quad (13)$$

where  $\text{MSE}_i^{\text{SV}}$  and  $\text{MSE}_i^{\text{SC}}$  are the simulated MSE values of SV and SC, respectively. In the upper panel of Figure 1, we present the scatter plots of  $\text{RD}_i$  in two scenarios, noting that the negative value of  $\text{RD}_i$  means the proposed SV method provides more accurate estimate than the SC method. From Figure 1, it is revealed that SV considerably improved the estimation accuracy over SC under scenario (I). On the other hand, it is natural result that SV is inefficient compared to SC under scenario (II) since SV uses



only local information for estimating hyperparameters. However, it should be pointed out that the difference between SV and SC is quite small in scenario (II) compared to the amount of improvement in scenario (I).

We next considered the Binomial-beta model:

$$z_i|\mu_i \sim \text{Bin}(n_i, \mu_i), \quad \mu_i \sim \text{Beta}(\nu(\mathbf{u}_i)m_i(\mathbf{u}_i), \nu(\mathbf{u}_i)(1 - m_i(\nu_i))), \quad i = 1, \dots, m, \quad (14)$$

where  $m = 60$ ,  $n_i = 20$  and  $m_i(\mathbf{u}_i) = \text{logistic}(\beta_0(\mathbf{u}_i) + \beta_1(\mathbf{u}_i)x_i)$ . We again considered the same two scenarios for hyperparameters. Based on  $R = 1000$  simulation runs, we computed the simulated MSE (12), and the percent relative differences (13) in two scenarios are presented in the lower panels of Figure 1. It is revealed that the proposed SV method works well in the binomial-beta model.

#### 4.2 Estimation error comparison in non-sampled area

We next investigated the estimation errors in non-sampled areas as discussed in Section 2.4. To this end, we considered 80 small areas in which the first  $m = 60$  areas are sampled areas and the last  $k = 20$  areas are non-sampled areas. Similarly to the previous section, we set  $n_i = 20$  for all sampled areas, and used the same data generating processes for the coordinates  $\mathbf{u}_i = (u_{1i}, u_{2i})$  and covariate  $x_i$ . Concerning the hyperparameters, we again considered the same two scenarios as used in Section 4.1. We first considered the Poisson-gamma model. The true means  $\mu_i$  were generated from the following model:

$$\mu_i \sim \Gamma(\nu(\mathbf{u}_i) \exp(\beta_0(\mathbf{u}_i) + \beta_1(\mathbf{u}_i)x_i), \nu(\mathbf{u}_i)), \quad i = 1, \dots, m + k.$$

On the other hand, the observations were generated only in the first  $m$  areas:

$$z_i|\mu_i \sim \text{Po}(n_i\mu_i), \quad i = 1, \dots, m.$$

Based on the observations  $z_1, \dots, z_m$ , we computed the estimates of  $\mu_{m+1}, \dots, \mu_{m+k}$  in non-sampled area, based on both spatially varying (SV) and spatially constant (SC) methods given in section 2.4. Based on  $R = 1000$  iterations, we simulated the area-level MSE defined in (12) in non-sampled area  $i = m + 1, \dots, m + k$ . To compare the results, we calculated the percentage relative difference (13) and present in Figure 2 in two scenarios. Moreover, we also considered the binomial-beta model:

$$\begin{aligned} z_i|\mu_i &\sim \text{Bin}(n_i, \mu_i), \quad i = 1, \dots, m. \\ \mu_i &\sim \text{Beta}(\nu(\mathbf{u}_i)m_i(\mathbf{u}_i), \nu(\mathbf{u}_i)(1 - m_i(\nu_i))), \quad i = 1, \dots, m + k, \end{aligned}$$

with  $m_i(\mathbf{u}_i) = \text{logistic}(\beta_0(\mathbf{u}_i) + \beta_1(\mathbf{u}_i)x_i)$ , and computed the percentage relative difference (13) between SV and SC methods in the same two scenarios, which are shown in Figure 2. From Figure 2, we can observe the similar results to the previous section, namely, SV works much better than SC in scenario (I) and SV produces inefficient estimates compared with SC in scenario (II) while the differences are quite small.

#### 4.3 Finite sample performances of MSE estimators

We finally investigate the finite sample performance of MSE estimator developed in Section 2.2. Similarly to the previous studies, we considered both Poisson-gamma and binomial-beta models. We assume there are  $m = 50$  areas divided into five groups with

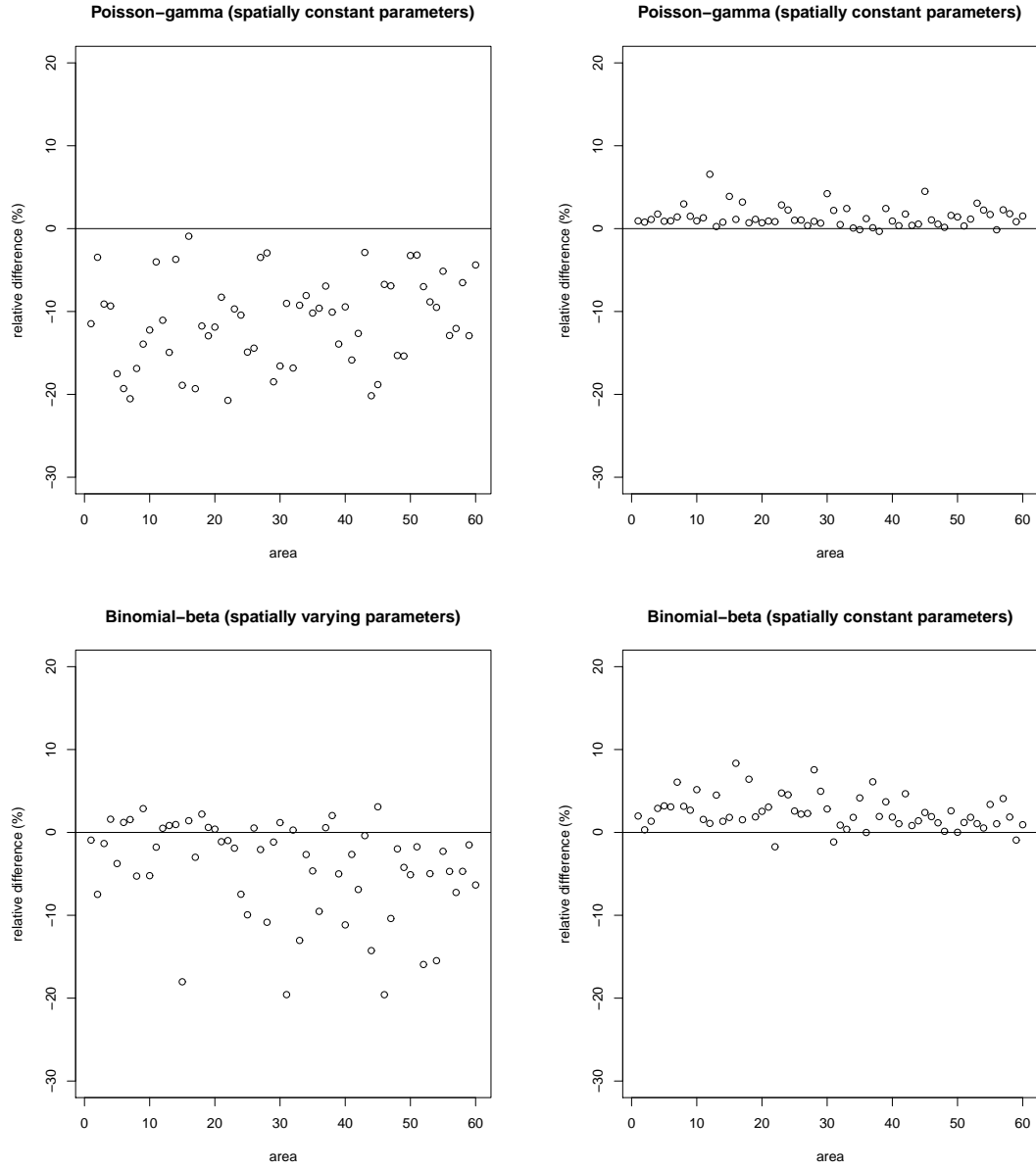


Figure 1: Ratios of the MSE (upper) and the bias (lower) of the predictors in the spatial and non-spatial binomial-beta model under a spatial non-stationary population (left) and a stationary population (right).

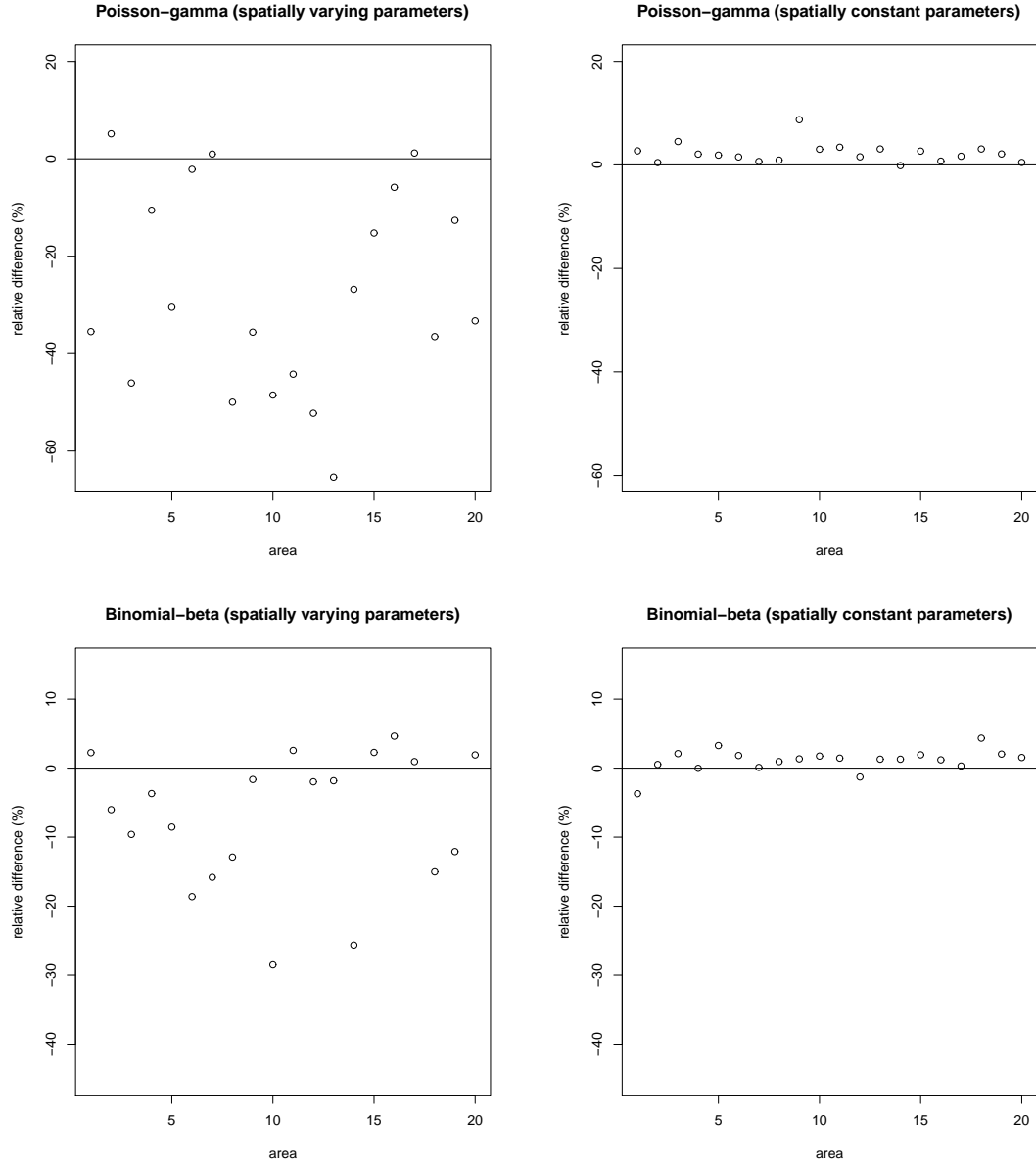


Figure 2: The percentage relative differences between the spatially varying method and spatially constant method in two scenarios in Poisson-gamma and binomial-beta models.

equal number of areas. The area-specific constant  $n_i$  are the same for areas within the same group. We used (10, 15, 20, 25, 30) for the group patterns of  $n_i$ .

The coordinate  $\mathbf{u}_i = (u_{1i}, u_{2i})$  were generated from the uniform distribution on  $(0, 1) \times (0, 1)$  and covariate  $x_i$  from the uniform distribution on  $(-1, 1)$ . In both Poisson-gamma (11) and binomial-beta (14) models, we set the true hyperparameters as  $\beta_0(\mathbf{u}_i) = u_{1i} - u_{2i} - 1$ ,  $\beta_1(\mathbf{u}_i) = \sqrt{u_{1i}^2 + u_{2i}^2}$  and  $\nu(\mathbf{u}_i) = 30 \exp(u_{1i} + u_{2i} - 1)$ . We first simulated the MSE (12) based on  $R = 100$  simulation runs, which are used as the true values of MSE in each area. For estimating these true MSEs, we used the bias-corrected MSE estimator given in (7) with  $B = 100$  bootstrap samples. For comparison, we also considered the naive estimator (6) which has considerable bias when  $m$  is not sufficiently large. Based on  $S = 100$  iterations, we calculated the percentage relative bias (RB) and coefficient of variation (CV), which are defined as

$$\text{RB}_i = \frac{1}{S} \sum_{s=1}^S \frac{\widehat{\text{MSE}}_i^{(s)} - \text{MSE}_i}{\text{MSE}_i}, \quad \text{CV}_i = \sqrt{\frac{1}{S} \sum_{s=1}^S \left( \frac{\widehat{\text{MSE}}_i^{(s)} - \text{MSE}_i}{\text{MSE}_i} \right)^2}$$

where  $\widehat{\text{MSE}}_i^{(s)}$  is the MSE estimate in the  $s$ th iteration and  $\text{MSE}_i$  is the true value. In Table 1, we report the averaged values of RB and CV within the same groups. It is noted that RB and CV of the naive MSE estimator are represented by RBN and CVN, respectively. From Table 1, it is revealed that the naive MSE estimator has the serious negative bias even when  $m = 50$ , which comes from the characteristics that the naive MSE estimator ignores the variability in estimating unknown hyperparameters. On the other hand, the bias-corrected MSE estimator provides relatively accurate estimates in terms of both RB and CV owing to the bootstrap bias correction.

Table 1: Percentage relative bias (RB) and coefficient of variation (CV) of the bias-corrected MSE estimator and the naive MSE estimator in Poisson-gamma and binomial-beta model. The RB and CV for the naive estimator are denoted by RBN and CVN, respectively.

|               | $n_i$ | 10     | 15     | 20     | 25     | 30     |
|---------------|-------|--------|--------|--------|--------|--------|
| Poisson-gamma | RB    | -6.30  | 16.37  | -3.44  | 7.58   | 1.68   |
|               | CV    | 43.87  | 45.26  | 40.41  | 36.49  | 32.14  |
|               | RBN   | -28.51 | -10.89 | -25.00 | -13.58 | -18.39 |
|               | CVN   | 49.97  | 45.23  | 48.28  | 43.94  | 38.99  |
| Binomial-beta | RB    | -0.84  | -5.03  | -6.24  | -0.55  | 2.05   |
|               | CV    | 47.37  | 39.93  | 38.41  | 36.45  | 32.35  |
|               | RBN   | -27.66 | -33.19 | -29.70 | -28.18 | -23.03 |
|               | CVN   | 48.27  | 47.53  | 46.14  | 43.71  | 37.67  |

## 5 Examples

### 5.1 Scottish lip cancer data

We first apply the proposed method to Scottish lip cancer data during the 6 years from 1975 to 1980 in each of the  $m = 56$  counties of Scotland, which was also analyzed

in Clayton and Kaldor (1987). For each county, the observed and expected number of cases are available, which are respectively denoted by  $z_i$  and  $n_i$ . Moreover, the proportion of the population employed in agriculture, fishing, or forestry is available for each county, thereby we used it as a covariate  $AFF_i$ , following Wakefield (2007). For each area,  $i = 1, \dots, m$ , we applied the spatially varying Poisson-gamma model:

$$z_i | \lambda_i \sim \text{Po}(n_i \lambda_i), \quad \lambda_i \sim \Gamma(\nu(\mathbf{u}_i) \exp(\beta_1(\mathbf{u}_i) + \beta_2(\mathbf{u}_i) AFF_i), \nu(\mathbf{u}_i)), \quad (15)$$

where  $\mathbf{u}_i = (u_{i1}, u_{i2})$ , and  $u_{i1}$  and  $u_{i2}$  are standardized longitude and latitude.

We first searched the optimal bandwidth by minimizing the criteria (5) and got  $b^* = 0.900$ . Then we computed the estimates of hyperparameters as well as the spatially varying empirical Bayes (SVEB) estimates of  $\lambda_i$  with  $b = b^*$ , which are shown in Figure 3. From Figure 3, it is observed that the hyperparameter estimates are dramatically changing from areas to areas. For comparison, we applied the conventional Poisson-gamma model:

$$z_i | \lambda_i \sim \text{Po}(n_i \lambda_i), \quad \lambda_i \sim \Gamma(\nu \exp(\beta_1 + \beta_2 AFF_i), \nu), \quad (16)$$

and the maximum likelihood estimates of the hyperparameters are  $\hat{\nu} = 2.13, \hat{\beta}_1 = -0.15$  and  $\hat{\beta}_2 = 5.18$ .

Let  $\hat{\lambda}_i^{\text{SVEB}}$  and  $\lambda_i^{\text{EB}}$  be the spatial varying empirical Bayes (SVEB) estimates from (15) and the empirical Bayes (EB) estimates from (16), respectively. In the left panel of Figure 4, we show the sample plot of the percentage relative difference  $100 \times (\hat{\lambda}_i^{\text{EB}} - \hat{\lambda}_i^{\text{SVEB}}) / \hat{\lambda}_i^{\text{SVEB}}$  against the log expected number of cases  $\log n_i$ . We can observe that the differences are larger in areas with small  $n_i$  while the difference gets smaller as  $n_i$  gets larger. This is because both SVEB and EB estimators are close to the direct estimator  $y_i$  in areas with large  $n_i$ . In the right panel of Figure 4, we present the sample plot of the square root of MSE (RMSE) estimates based on 500 bootstrap samples against  $\log n_i$ , which reveals that the RMSE decreases as  $n_i$  increases.

Finally, we computed the benchmarked estimator  $\hat{\lambda}_i^{\text{C}}$  from (8) with the weight  $c_i = n_i / \sum_{k=1}^m n_k$ , and the relative differences to  $\hat{\lambda}_i = \hat{\lambda}_i^{\text{SVEB}}$  are presented in the left panel of Figure 5. The figure shows that the differences are increasing with respect to  $n_i$ , which comes from the choice of the benchmarking weight  $c_i$ . However, in most areas, the relative differences are smaller than 2%, so that  $\hat{\lambda}_i^{\text{C}}$  and  $\hat{\lambda}_i$  is quite similar. Based on 500 bootstrap replications, we calculated the excess MSE estimates by using (9) and computed the ratio to the MSE estimates of SVEB. The histogram of the ratio is given in the right panel of Figure 5, which shows that the percentage of the risk inflation is at most 1.4%.

## 5.2 Spanish Poverty rates data

We next use synthetic income data set in Spanish provinces, which is available in the R package `sae`. In the data set, unit data are available for 52 areas. Let  $N_i, i = 1, \dots, m$  denote the population sizes of the areas. Let  $E_{ij}$  be the equivalized disposable income calculated following the standard procedure of the Spanish Statistical Institute, and  $z$  be the poverty line. The poverty rate for area  $i$  is defined as  $p_i = N_i^{-1} \sum_{j=1}^{N_i} I(E_{ij} < z)$ . Unfortunately, we do not observe all  $E_{ij}$ 's but observe only  $E_{ij}, j = 1, \dots, n_i$ . A direct estimator  $y_i$  of  $p_i$  is given by

$$y_i = \frac{1}{n_i} \sum_{j=1}^{n_i} I(E_{ij} < z),$$

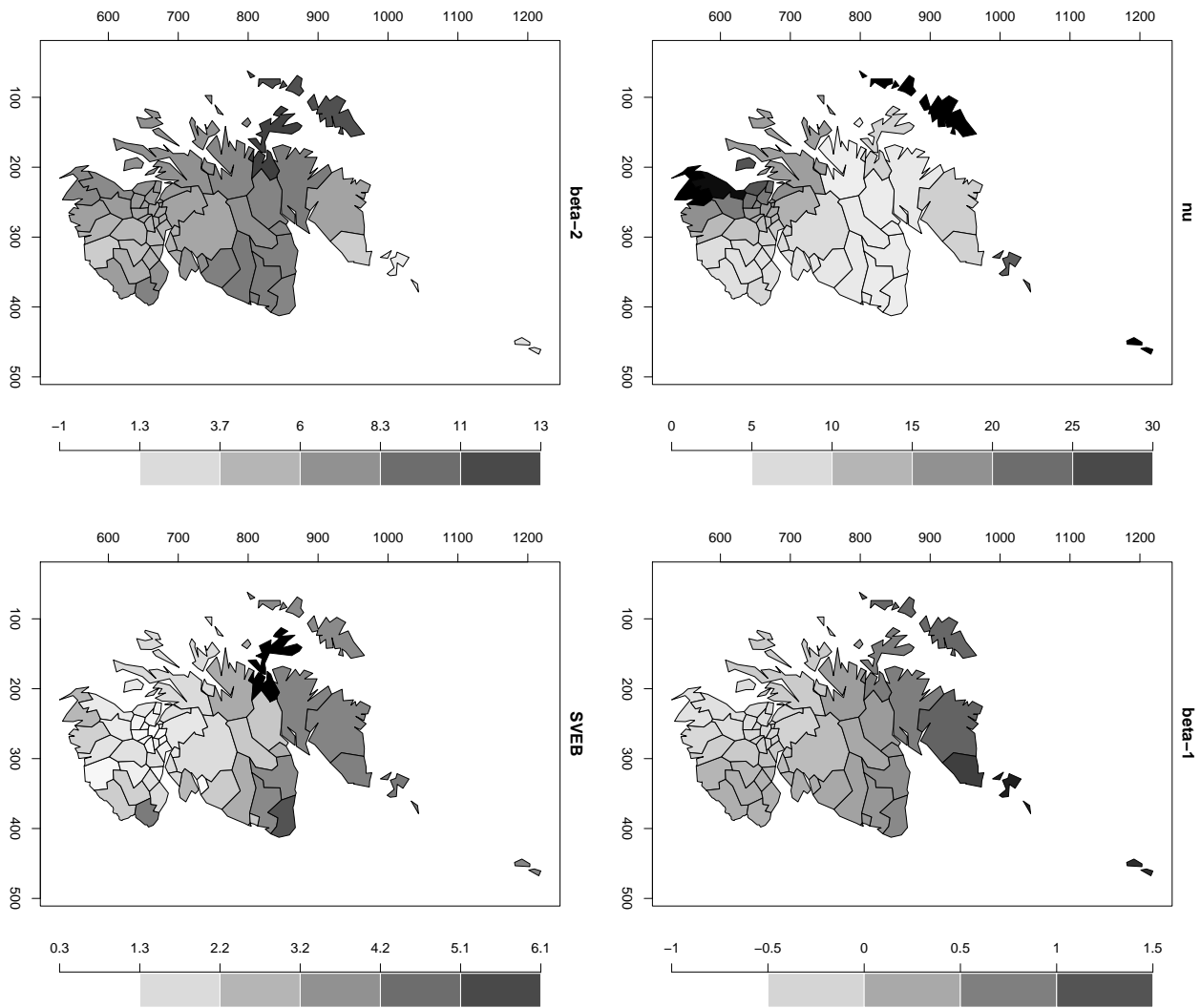


Figure 3: Spatial distributions of  $\hat{\nu}(\mathbf{u}_i)$  (top-left),  $\hat{\beta}_1(\mathbf{u}_i)$  (top-right),  $\hat{\beta}_1(\mathbf{u}_i)$  (bottom-left) and  $\hat{\lambda}_i$  (bottom-right).

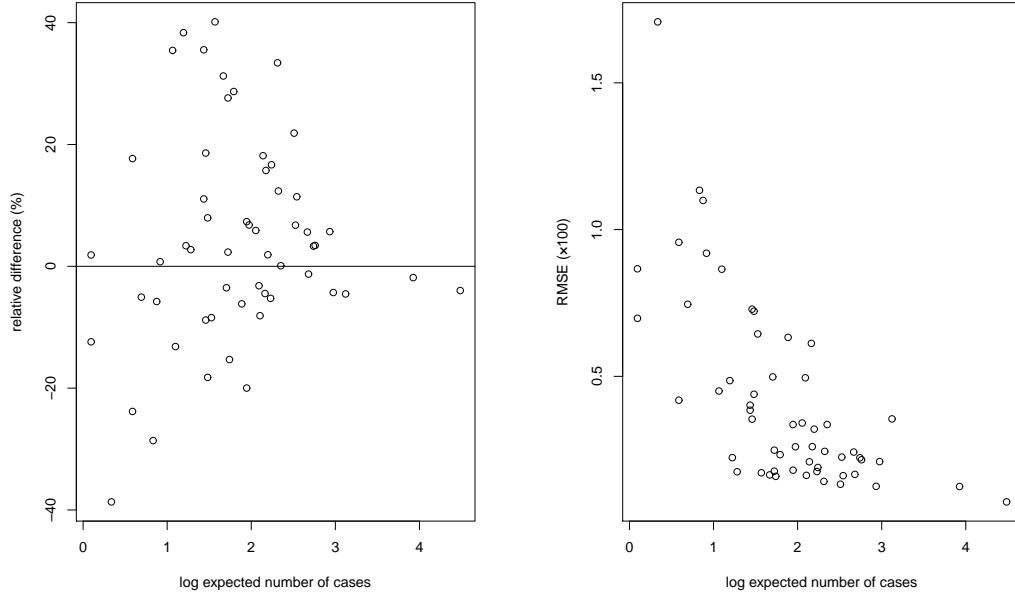


Figure 4: Sample plots of the percentage relative difference between SVEB and EB estimates:  $100 \times (\hat{\lambda}_i^{\text{EB}} - \hat{\lambda}_i^{\text{SVEB}}) / \hat{\lambda}_i^{\text{SVEB}}$  (left) and the ratio of the squared root of MSE (RMSE) estimates of SVEB to those of EB (right) in Scottish lip cancer data.

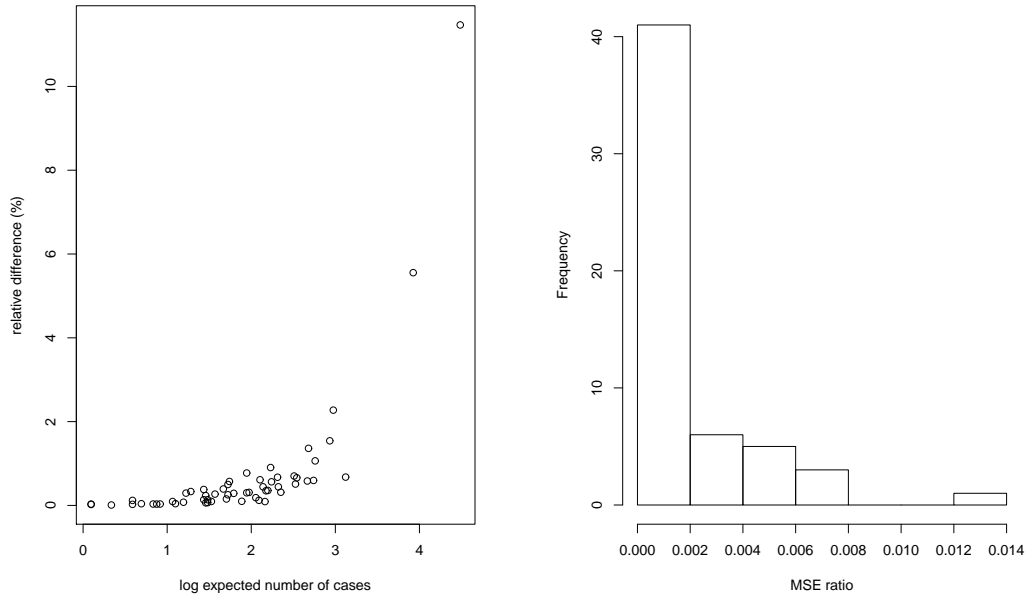


Figure 5: Sample plots of the percentage relative difference between SVEB and the benchmarked estimator:  $100 \times (\hat{\lambda}_i^{\text{C}} - \hat{\lambda}_i) / \hat{\lambda}_i$  (left) and histogram of the excess MSE estimates (right) in Scottish lip cancer data.



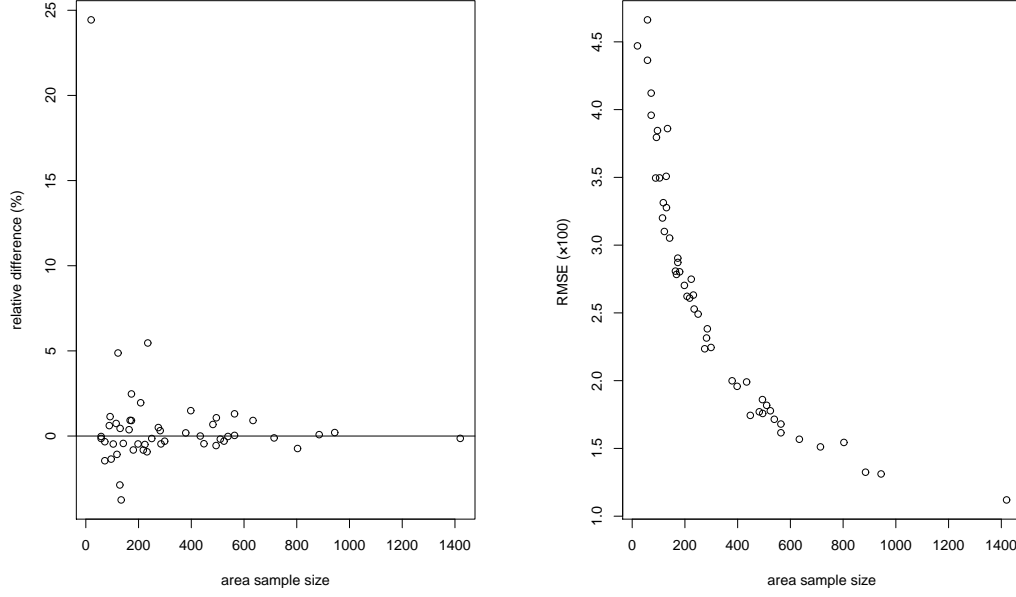


Figure 6: Sample plots of the percentage relative difference between SVEB and EB estimates:  $100 \times (\hat{p}_i^{\text{EB}} - \hat{p}_i^{\text{SVEB}}) / \hat{p}_i^{\text{SVEB}}$  (left) and the ratio of the squared root of MSE (RMSE) estimates of SVEB to those of EB (right) in Spanish poverty rate data.

where we set  $z$  as 0.6 times median of all the observed income  $E_{ij}$ 's, following Molina and Rao (2010). As area-level covariates, we used area-level rates of female and labors, which are respectfully denoted by  $\text{fe}_i$  and  $\text{lab}_i$ . Since two provinces, PalmasLas and Tenerife, are very far away from the other provinces, we have omitted in this study. Then, we applied the following binomial-beta model for  $i = 1, \dots, m$ :

$$y_i | p_i \sim \text{Bin}(n_i, p_i), \quad p_i \sim \text{Beta}(\nu(\mathbf{u}_i)m_i(\mathbf{u}_i), \nu(\mathbf{u}_i)(1 - m_i(\mathbf{u}_i))), \quad (17)$$

where  $m_i(\mathbf{u}_i) = \text{logistic}(\beta_1(\mathbf{u}_i) + \beta_2(\mathbf{u}_i)\text{fe}_i + \beta_3(\mathbf{u}_i)\text{lab}_i)$  and  $\mathbf{u}_i = (u_{i1}, u_{i2})$ , and  $u_{i1}$  and  $u_{i2}$  are standardized longitude and latitude. For comparison, we also applied the conventional binomial-beta model:

$$y_i | p_i \sim \text{Bin}(n_i, p_i), \quad p_i \sim \text{Beta}(\nu m_i, \nu(1 - m_i)), \quad (18)$$

with  $m_i = \text{logistic}(\beta_1 + \beta_2\text{fe}_i + \beta_3\text{lab}_i)$ .

We found that the optimal bandwidth is  $b^* = 2.42$ , and some empirical quintiles of the hyperparameter estimates are provided in Table 2. From Table 2, we can observe that the medians of spatially varying hyperparameter estimates in spatially varying model (17) is close to the point estimates in the conventional model (18). In the left panel in Figure 6, we presented the percentage relative difference  $100 \times (\hat{p}_i^{\text{EB}} - \hat{p}_i^{\text{SVEB}}) / \hat{p}_i^{\text{SVEB}}$ , where  $\hat{p}_i^{\text{SVEB}}$  and  $\hat{p}_i^{\text{EB}}$  are empirical Bayes estimates from (17) and (18), respectively. We can observe that the differences are smaller than 6% in all the areas except for the one area, and the differences vanishes as the area sample size  $n_i$  gets large. Based on 500 bootstrap replications, we computed the MSE estimates of  $\hat{p}_i^{\text{SVEB}}$ , and the square root of MSE (RMSE) estimates are given in the right panel of Figure 6, which shows the natural result that the MSE decreases with respect to  $n_i$ .

We finally computed the benchmarked estimator of  $p_i$  in model (17), and we found that the percentage relative difference between the SVEB and benchmarked estimates are smaller than 0.15% and the excess risks in benchmarking based on 500 bootstrap samples were negligibly small.

Table 2: Quantiles of hyperparameter estimates in spatially varying (SV) model and point estimates in spatially constant (SC) model in Spanish poverty rate data.

|                 | SV    |       |       |       |        | SC       |
|-----------------|-------|-------|-------|-------|--------|----------|
|                 | 0%    | 25%   | 50%   | 75%   | 100%   | Estimate |
| $\hat{\beta}_1$ | -7.96 | -3.34 | -2.28 | -1.03 | 1.11   | -2.70    |
| $\hat{\beta}_2$ | -3.47 | 2.13  | 3.50  | 5.26  | 10.65  | 3.85     |
| $\hat{\beta}_3$ | -4.54 | -2.16 | -1.69 | -0.90 | 3.22   | -1.19    |
| $\hat{\nu}$     | 42.33 | 44.12 | 48.06 | 51.59 | 103.06 | 46.32    |

## 6 Discussion

We have developed spatially varying empirical Bayes methods based on the local likelihood estimation in which the optimal bandwidth in a kernel function is determined by cross validation. The model we considered can be regarded as a generalization of the two-stage hierarchical area-level models based on natural exponential family, proposed by Ghosh and Maiti (2004). The model includes Fay-Herriot model, Poisson-gamma model, and binomial-beta model as special cases, so that the model is applicable for both continuous, count and binary data. We have considered some problems including the MSE estimation, benchmarking estimation, estimating in non-sampled areas. The proposed methods were compared with the conventional non-spatial models through simulation and empirical studies, and we have found that the proposed method works well and would improve the estimation accuracy of the traditional methods.

The possible drawback of the proposed method is its computational costs when the number of areas  $m$  is large. For specified bandwidth  $b$ , it requires  $m$  times maximization of the weighted log-marginal likelihood (4) to compute hyperparameter estimates in each area, thereby its computational cost increases linearly depending on  $m$ . A possible solution is to assume that  $m$  areas can be classified in  $G$  groups, where  $G$  is much smaller than  $m$ , and the hyperparameters are the same in all the areas within the same group. This can reduce the number of maximization from  $m$  to  $G$  for each  $b$ . However, it is not straightforward how we can efficiently divide the areas. Hence, the detailed consideration about the issue seems to exceed the scope of this paper and we left the problem as a valuable future study.

**Acknowledgments.** We appreciate valuable comments from Tatsuya Kubokawa and Hisashi Noma. This work was supported by JSPS KAKENHI Grant Numbers, JP16H07406, JP16K17101 and JP16K17153.

## References

- [1] Bandyopadhyay, D., Reich, B. J. and Slate, E. H. (2009). Bayesian modeling of multivariate spatial binary data with applications to dental caries. *Statistics in Medicine*, **28**, 3492-3508.
- [2] Bell, W.R., Datta, G.S. and Ghosh, M. (2013). Benchmarking small area estimators. *Biometrika*, **100**, 189-202.
- [3] Brent, R. (1973). *Algorithms for Minimization without Derivatives*. Englewood Cliffs N.J. Prentice-Hall.
- [4] Brunsdon, C., Fotheringham, A.S., and Charlton, M. (1998). Geographically weighted regression—modelling spatial non-stationarity. *Journal of the Royal Statistical Society: Series D*, **47**, 431-443.
- [5] Butar, F. B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *Journal of Statistical Planning and Inference*, **112**, 63-76.
- [6] Chambers, R., Dreassi, E. and Salvati, N. (2014). Disease mapping via negative binomial regression M-quantiles. *Statistics in Medicine*, **33**, 4805-4824.
- [7] Chandra, H., Salvati, N., Chambers, R. and Tzavidis, N. (2012). Small area estimation under spatial nonstationarity. *Computational Statistics & Data Analysis*, **56**, 2875–2888.
- [8] Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671-681.
- [9] Datta, G.S., Ghosh, M., Steorts, R. and Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, **20**, 574-588.
- [10] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 183-196.
- [11] Dempster, A., Laird, N., and Rubin, D. (1977). Maximum Likelihood From Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society: Series B*, **39**, 1-38.
- [12] Fay, R. and Herriot, R. (1979). Estimators of income for small area places: An application of James-Stein procedures to census. *Journal of the American Statistical Association*, **74**, 341-353.
- [13] Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002). *Geographically weighted regression*. Wiley, West Sussex.
- [14] Ghosh, M. (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association*, **87**, 533-540.
- [15] Ghosh, M. and Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, **91**, 95-112.
- [16] Hall, P. and Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society: Series B*, **68**, 221-238.
- [17] Jiang, J. (2006). *Linear and generalized linear mixed models and their applications*, Springer.
- [18] Kubokawa, T., Hasukawa, M. and Takahashi, K. (2014). On measuring uncertainty of benchmarked predictors with application to disease risk estimate. *Scandinavian Journal of Statistics*, **35**, 394-413.

- [19] Marhuenda, Y., Molina, I. and Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics & Data Analysis*, **58**, 308-325.
- [20] Molina, I. and Rao, J. N. K. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics*, **38**, 369-385.
- [21] McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. Wiley, New York.
- [22] Morris, C. (1982). Natural exponential families with quadratic variance functions. *The Annals of Statistics*, **10**, 65-80.
- [23] Morris, C. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *The Annals of Statistics*, **11**, 515-529.
- [24] Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*, **28**, 40-68.
- [25] Prasad, N. and Rao, J. N. K. (1990). The estimation of mean-squared errors of small-area estimators. *Journal of the American Statistical Association*, **90**, 758-766.
- [26] Rao, J.N.K. and Molina, I. (2015) *Small Area Estimation, 2nd Edition*. Wiley.
- [27] Salvati, N., Tzavidis, N., Pratesi, M. and Chambers, R. (2012). Small area estimation via m-quantile geographically weighted regression. *Test*, **21**, 1-28.
- [28] Silva, A. R. and Rodrigues, T. C. V. (2014). Geographically weighted negative binomial regression—incorporating overdispersion. *Statistics and Computing*, **24**, 769-783.
- [29] Steorts, R. and Ghosh, M. (2013). On estimation of mean squared errors of benchmarked empirical Bayes estimators. *Statistica Sinica*, **23**, 749-767.
- [30] Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, **82**, 559-567.
- [31] Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, **8**, 158-183.
- [32] Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, **31**, 949-952.